

DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis

Year: 2020 | Citations: 264 | Authors: Ming Tao, Hao Tang, Fei Wu, Xiaoyuan Jing, Bingkun Bao

Abstract

Synthesizing high-quality realistic images from text descriptions is a challenging task. Existing text-to-image Generative Adversarial Networks generally employ a stacked architecture as the backbone yet still remain three flaws. First, the stacked architecture introduces the entanglements between generators of different image scales. Second, existing studies prefer to apply and fix extra networks in adversarial learning for text-image semantic consistency, which limits the supervision capability of these networks. Third, the cross-modal attention-based text-image fusion that widely adopted by previous works is limited on several special image scales because of the computational cost. To these ends, we propose a simpler but more effective Deep Fusion Generative Adversarial Networks (DF-GAN). To be specific, we propose: (i) a novel one-stage text-to-image backbone that directly synthesizes high-resolution images without entanglements between different generators, (ii) a novel Target-Aware Discriminator composed of Matching-Aware Gradient Penalty and One-Way Output, which enhances the text-image semantic consistency without introducing extra networks, (iii) a novel deep text-image fusion block, which deepens the fusion process to make a full fusion between text and visual features. Compared with current state-of-the-art methods, our proposed DF-GAN is simpler but more efficient to synthesize realistic and text-matching images and achieves better performance on widely used datasets. Code is available at <https://github.com/tobran/DF-GAN>.